Qualitative robustness of statistical functionals under strong mixing

Henryk Zähle[*]

Abstract

A new concept of (asymptotic) qualitative robustness for plug-in estimators based on identically distributed possibly *dependent* observations is introduced, and it is shown that Hampel's theorem for general metrics d still holds. Since Hampel's theorem assumes the UGC property w.r.t. d, i.e. convergence in probability of the empirical probability measure to the true marginal distribution w.r.t. d uniformly in the class of all admissible laws on the sample path space, this property is shown for a large class of strongly mixing laws for three different metrics d. For real-valued observations the UGC property is established for both the Kolomogorov ϕ-metric and the Lévy ψ-metric, and for observations in a general locally compact and second countable Hausdorff space the UGC property is established for a certain metric generating the ψ-weak topology. The key is a new uniform weak LLN for strongly mixing random variables. The latter is of independent interest and relies on Rio's maximal inequality.

Keywords: plug-in estimator, qualitative robustness, Hampel's theorem, strong mixing, Rio's maximal inequality, function bracket, Kolmogorov ϕ-metric, locally compact and second countable Hausdorff space, ψ-weak topology, Lévy ψ-metric, uniform Glivenko-Cantelli theorem, uniform weak law of large numbers

---

[*]Department of Mathematics, Saarland University, Germany, `zaehle@math.uni-sb.de`

1. Introduction

Let $\mathcal{M}$ be a class of probability measures on some measurable space E, and T be a mapping (statistical functional) from $\mathcal{M}$ into a measurable space $\mathbf{T}$. Let $(X_i)_{i \in \mathbb{N}}$ be a sequence of E-valued random elements (observations) being identically distributed according to $\mu \in \mathcal{M}$. If $\widehat{m}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ denotes the empirical distribution of $X_1, \ldots, X_n$, then $T(\widehat{m}_n)$ can provide a reasonable estimator for $T(\mu)$. Informally, the sequence $(T(\widehat{m}_n))$ is qualitatively robust when for large n a small change in μ results only in a small change of the law of the estimator $T(\widehat{m}_n)$. More precisely, given a subset $\mathcal{P}_1 \subset \mathcal{M}$, the sequence of estimators $(T(\widehat{m}_n))$ is said to be qualitatively $\mathcal{P}_1$-robust at $\mu \in \mathcal{P}_1$ if for every $\varepsilon > 0$ there are some $\delta > 0$ and $n_0 \in \mathbb{N}$ such that

$$\nu \in \mathcal{P}_1, \quad d(\mu, \nu) \leq \delta \quad \Longrightarrow \quad d'(\text{law}\{T(\widehat{m}_n)|\mu\}, \text{law}\{T(\widehat{m}_n)|\nu\}) \leq \varepsilon \quad \forall\, n \geq n_0, \quad (1)$$

where d and d' are metrics on $\mathcal{P}_1$ and on the class of all probability measures on $\mathbf{T}$, respectively. The basic concept of qualitative robustness was initiated by Hampel [16, 17], but the above version of qualitative robustness is due to Huber [19]. Huber's version is also called asymptotic robustness (cf. [25]) and differs from the original definition in [16, 17] in that the right-hand side in (1) is not required to hold for all $n \in \mathbb{N}$ but only for all $n \geq n_0$ for some $n_0 = n_0(\varepsilon)$. For background see also [11, 18, 20, 21, 24, 27] and references therein.

The definition of qualitative robustness stated above was introduced by Hampel and Huber to capture the case of independent observations. To capture also the case of dependent observations, various authors departed from this definition and considered, instead of a metric d on a class of probability measures on E, a metric $d_{\mathbb{N}}$ on a class of probability measures on the infinite product space $E^{\mathbb{N}}$ or a sequence of pseudo-metrics (d_n) on classes of probability measures on E^n, $n \in \mathbb{N}$; cf. [3, 7, 10, 17, 25]. However, in the usual situation where one is interested in the estimation of an aspect $T(\mu)$ of the marginal distribution μ based on the observations $X_1, \ldots, X_n$, and where the contaminated observations are still identically distributed (according to some ν close to μ), it might be also reasonable to retain the original definition. After all, for increasing sample size n the impact of the data dependence on the estimation often declines. So one might hope that if the dependence structures induced by the "admissible" probability measures on $E^{\mathbb{N}}$ (which play the role of the laws of the sequences of identically distributed observations) are subject to a common constraint, then the implication (1) still holds for the class $\mathcal{P}_1$ of the marginal distributions. In other words, under a common constraint for the dependence structures, it might be sufficient to ensure a small distance between the marginal distributions μ and ν in order to obtain a small distance between the distributions of the plug-in estimators, based on large n, under two "admissible" laws on $E^{\mathbb{N}}$ with marginal distributions μ and ν, respectively.

In this article, we will demonstrate that the latter is in fact true under fairly weak

constraints for the dependence structures. In Section 2, we adapt Huber's definition of qualitative robustness to the case of dependent observations and establish the analogue of Hampel's theorem. Since the latter assumes the UGC property, i.e. convergence in probability of the empirical probability measure to the true marginal distribution uniformly in the class of all "admissible" laws on $E^{\mathbb{N}}$ (cf. Definition 2.3 below), this property will be established for a large class of strongly mixing laws on $E^{\mathbb{N}}$ for three different metrics. In Section 3.1, we consider real-valued observations ($E = \mathbb{R}$) and verify the UGC property for both the Kolomogorov ϕ-metric and the Lévy metric. In Section 3.2, we assume observations in a general locally compact and second countable Hausdorff space E and verify the UGC property for a certain metric generating the ψ-weak topology. For both examples the key is a new uniform weak LLN for strongly mixing random variables which is given in the Appendix A and which relies on Rio's maximal inequality.

It should be stressed that for the considerations of this article it is essential that the definition of qualitative robustness (Definition 2.1 below) is in line with Huber's version of qualitative robustness [19], i.e. with asymptotic robustness. Indeed, from Example 1.18 in [5] it is easily seen that for *fixed n*, weak dependence can change the distribution of an estimator even if the marginal distributions of the observed data are the same. In this respect, the intension of this article differs from the objective of the existing literature on qualitative robustness for dependent observations [3, 7, 10, 17, 25] where the estimators are demanded to be "stable" not only for large but also for small samples (and are allowed to be more general than plug-in estimators). The latter notion of robustness requires a more sophisticated definition compared to Definition 2.1 below. See, for instance, [3] for an informative discussion on a proper choice for the definition of qualitative robustness in this context.

2. Qualitative robustness and a Hampel theorem

Let $(E, \mathcal{E})$ be a measurable space, $\Omega := E^{\mathbb{N}}$, $\mathcal{F} := \mathcal{E}^{\mathbb{N}}$, X_i be the $i$the coordinate projections on Ω, and $\mathbb{P}_i := \mathbb{P} \circ X_i^{-1}$ be the ith marginal distribution of a probability measures $\mathbb{P}$ on $(\Omega, \mathcal{F})$. Let $\mathcal{P}$ be a class of probability measures on $(\Omega, \mathcal{F})$ such that $\mathbb{P}_1 = \mathbb{P}_2 = \cdots$ for every $\mathbb{P} \in \mathcal{P}$. Let $\mathcal{P}_1 := \{\mathbb{P}_1 : \mathbb{P} \in \mathcal{P}\}$ be the corresponding class of all marginal distributions, and $\mathcal{M}$ be any subset of the class $\mathcal{M}_1(E)$ of all probability measures on $(E, \mathcal{E})$ such that $\mathcal{P}_1 \subset \mathcal{M}$. Let $(\mathbf{T}, \mathcal{T})$ be a measurable space, and $T : \mathcal{M} \to \mathbf{T}$ be a mapping (statistical functional). For every $n \in \mathbb{N}$, we assume that the mapping

$$\widehat{T}_n(x) = \widehat{T}_n(x^{(n)}) := T(\widehat{m}_{x^{(n)}}), \qquad x = (x_1, x_2, \ldots) \in \Omega \tag{2}$$

is $(\mathcal{E}^{\mathbb{N}}, \mathcal{T})$-measurable, where $\widehat{m}_{x^{(n)}} := \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}$ denotes the empirical probability measures associated with $x^{(n)} := (x_1, \ldots, x_n)$. For (2) to be well defined, we assume that the set of all such empirical probability measures is contained in $\mathcal{M}$. Notice that $\widehat{T}_n$ provides

an estimator for $T(\mu)$. We let d' be some metric on the set $\mathcal{M}_1(\mathbf{T})$ of all probability measures on $(\mathbf{T}, \mathcal{T})$, and d be some metric on $\mathcal{P}_1$.

Definition 2.1 (Qualitative robustness) *Let us take the notation from above, and let* $\mathbb{P} \in \mathcal{P}$. *Then the sequence* $(\widehat{T}_n)$ *of estimators is said to be* qualitatively $\mathcal{P}$-marginally robust *at* $\mathbb{P}$ *w.r.t.* (d, d') *if for every* $\varepsilon > 0$ *there are some* $\delta > 0$ *and* $n_0 \in \mathbb{N}$ *such that*

$$\mathbb{Q} \in \mathcal{P}, \quad d(\mathbb{P}_1, \mathbb{Q}_1) \le \delta \qquad \Longrightarrow \qquad d'(\mathbb{P} \circ \widehat{T}_n^{-1}, \mathbb{Q} \circ \widehat{T}_n^{-1}) \le \varepsilon \quad \forall n \ge n_0.$$

Sometimes also the functional T itself will be called qualitatively $\mathcal{P}$-marginally robust at $\mathbb{P}$ if the corresponding sequence of plug-in estimators $(\widehat{T}_n)$ is.

Remark 2.2 If every $\mathbb{P} \in \mathcal{P}$ is an infinite product measure, i.e. $\mathbb{P} = \mathbb{P}_1^{\mathbb{N}}$, then Definition 2.1 coincides with the classical definition of qualitative robustness for independent observations; cf. [19, 20, 21]. $\diamond$

In applications, the validity of qualitative robustness in the sense of Definition 2.1 is typically hard to check "directly". So it is natural to ask for transparent sufficient conditions. In the framework of Remark 2.2 the celebrated Hampel theorem provides a sufficient condition for qualitative robustness when $\mathcal{M} = \mathcal{P}_1 = \mathcal{M}_1(E)$, E is Polish, and d and d' are the Prohorov metrics; cf. [11, 17, 19, 20, 24]. In [21], this result was extended to more general metric spaces $(\mathcal{P}_1, d)$ but still in the framework of Remark 2.2. In Theorem 2.4 below we will formulate a version of Hampel's criterion which can also be applied to our general setting. As usual, we choose d' as the Prohorov metric. To this end, we assume that $\mathbf{T}$ is equipped with a complete and separable metric $d_{\mathbf{T}}$ and that $\mathcal{T}$ is the corresponding Borel σ-field. Recall that the Prohorov metric is given by

$$d'_{\mathrm{Proh}}(\mu, \nu) := \inf\{\varepsilon > 0 : \mu[A] \le \nu[A^\varepsilon] + \varepsilon \text{ for all } A \in \mathcal{T}\},$$

where $A^\varepsilon := \{t \in T : \inf_{a \in A} d_{\mathbf{T}}(t, a) \le \varepsilon\}$ is the ε-hull of A. Moreover, we set $\widehat{m}_n(x) := \widehat{m}_{x^{(n)}}$ for all $x \in \Omega$. In the following definition, which is a generalization of Definition 2.3 in [21], the acronym UGC stands for "uniform (weak) Glivenko–Cantelli".

Definition 2.3 (UGC property) *We say that* $\mathcal{P}$ *admits the UGC property w.r.t.* d *if for every* $\delta > 0$

$$\lim_{n \to \infty} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[d(\widehat{m}_n, \mathbb{P}_1) \ge \delta] = 0. \tag{3}$$

Theorem 2.4 (Hampel-type theorem) *Assume that* $\mathcal{P}$ *admits the UGC property w.r.t.* d, *and let* $\mathbb{P} \in \mathcal{P}$. *Then, if the mapping* T *is continuous at* $\mathbb{P}_1$ *w.r.t.* $(d, d_{\mathbf{T}})$, *the sequence* $(\widehat{T}_n)$ *is qualitatively* $\mathcal{P}$-*marginally robust at* $\mathbb{P}$ *w.r.t.* (d, d'_{Proh}).

The proof of Theorem 2.4 can be found in Section 4.

Remark 2.5 Let $\mathcal{M}_{1,\mathrm{emp}}$ denote the space of all empirical probability measures $\widehat{m}_{x^{(n)}}$ with $x \in E^{\mathbb{N}}$ and $n \in \mathbb{N}$, and recall that we assumed $\mathcal{M}_{1,\mathrm{emp}} \subset \mathcal{M}$. It is clear from the proof in Section 4 that in Theorem 2.4 it suffices to require that T is $\mathcal{M}_{1,\mathrm{emp}}$-continuous at $\mathbb{P}_1$ w.r.t. $(d, d_{\mathbf{T}})$, meaning that for every $\varepsilon > 0$ there is some $\delta > 0$ such that for all $\nu \in \mathcal{M}_{1,\mathrm{emp}}$ with $d(\mathbb{P}_1, \nu) \leq \delta$ we have that $d_{\mathbf{T}}(T(\mathbb{P}_1), T(\nu)) \leq \varepsilon$. $\diamond$

Remark 2.6 Adapting the proof of Theorem 2.6 in [21], one also obtains a sort of converse of Hampel's theorem. More precisely, fix $\mathbb{P} \in \mathcal{P}$ and assume that $(\widehat{T}_n)$ is weakly consistent w.r.t. $d_{\mathbf{T}}$ (i.e. $\widehat{T}_n$ converges in $\mathbb{Q}$-probability to $T(\mathbb{Q}_1)$ w.r.t. $d_{\mathbf{T}}$) at every $\mathbb{Q} \in \mathcal{P}$ for which $\mathbb{Q}_1$ lies in some neighborhood of $\mathbb{P}_1$ w.r.t. d. Then qualitative $\mathcal{P}$-marginal robustness of the sequence $(\widehat{T}_n)$ at $\mathbb{P}$ w.r.t. (d, d'_{Proh}) implies that the mapping T is $\mathcal{P}_1$-continuous at $\mathbb{P}_1$ w.r.t. $(d, d_{\mathbf{T}})$. The latter means that for every $\varepsilon > 0$ there is some $\delta > 0$ such that $d_{\mathbf{T}}(T(\mathbb{P}_1), T(\nu)) \leq \varepsilon$ for every $\nu \in \mathcal{P}_1$ with $d(\mathbb{P}_1, \nu) \leq \delta$. $\diamond$

In Section 3, we will give two examples for classes of probability measures on $(\Omega, \mathcal{F})$ admitting the UGC property. To motivate the Hampel-type Theorem 2.4, we will also discuss continuity of particular statistical functionals w.r.t. the involved metrics.

The classical choice for d in the framework of Definition 2.1 is any metric generating the weak topology. The most prominent examples are the Prohorov metric and the Lévy metric used in [11, 17, 19, 20, 24] and many further references. Another example is the bounded Lipschitz metric used, for instance, in [8, 12, 19, 20]. However, for some purposes it is somewhat restrictive to use exclusively a metric generating the weak topology. The use of such a metric creates a sharp division of the class of statistical functionals T into those for which $(\widehat{T}_n)$ is "robust" and those for which $(\widehat{T}_n)$ is "not robust". Indeed, Hampel's theorem says that $(\widehat{T}_n)$ is "robust" if and only if T is continuous w.r.t. the weak topology. But the distributions of the plug-in estimators of two statistical functionals being not continuous w.r.t. the weak topology may react quite different to changes in the underlying (marginal) distribution, just as these plug-in estimators may have quite different influence functions. For this reason it was proposed in [21] to investigate $(\widehat{T}_n)$ for qualitative robustness w.r.t. more general metrics, where it is clear that qualitative robustness w.r.t. a metric d_1 is a stronger condition than qualitative robustness w.r.t. a metric $d_2 \leq d_1$. In particular, a statistical functional T_1 can be considered to have a higher "degree of robustness" than another statistical functional T_2 when T_1 is qualitatively robust for any choice of d for which T_2 is qualitatively robust. In this way, it gets possible to differentiate plug-in estimators w.r.t. qualitative robustness within the class of statistical functionals that are not weakly continuous. Sensible classes of metrics that can be studied in this context are, for instance, $\{d_{(\phi)}\}_\phi$ and $\{d_{\psi,\mathrm{vag}}\}_\psi$ to be introduced in (4) and (7), respectively, where ϕ and ψ (or rather their increases) can be seen as gauges for the strictness of the metric and thus for the "degree of robustness". For

details see [21, 22]. The latter reference provides in particular a rigorous quantification of the "degree of robustness" for convex risk functionals.

The preceding discussion counters somewhat the conventional point of view that the sequence $(\widehat{T}_n)$ of plug-in estimators can be considered to be qualitatively robust exclusively when T is continuous w.r.t. the weak topology. But even if one insists on the use of the weak topology, the considerations below provide new results. For instance, Corollary 3.7 and Theorem 2.4 together yield a nontrivial generalization of the classical Hampel theorem in the form of [19, Theorem 2.21]. In this theorem the underlying metric is the Lévy metric which generates the weak topology.

In the sequel we will repeatedly work with left- and rightcontinuous inverses. Recall that the leftcontinuous inverse of any nondecreasing function $H : \mathbb{R} \to \mathbb{R}$ is defined by $H^\leftarrow(t) := \inf\{y \in \mathbb{R} : H(y) \geq t\}$ with the convention $\inf \emptyset = \infty$. The rightcontinuous inverse $H^\rightarrow$ of any nonincreasing function $H : \mathbb{R}_+ \to \mathbb{R}_+$ is defined by $H^\rightarrow(t) := \sup\{y \in \mathbb{R}_+ : H(y) > t\}$ with the convention $\sup \emptyset := 0$. We will also repeatedly use the notion of strong mixing which is recalled in the Appendix A. The nth strong mixing coefficient of the coordinate process (X_i) under the law $\mathbb{P}$ will be denoted by $\alpha_\mathbb{P}(n)$.

3. Examples

3.1. Strong mixing, and Kolmogorov ϕ-metric or Lévy metric

In this section, we will see that in the case $E = \mathbb{R}$ a large class of probability measures on $(\Omega, \mathcal{F})$ admits the UGC property w.r.t. (a weighted version of) the Kolmogorov metric. As a corollary we will also obtain the UGC property w.r.t. the Lévy metric. Let ϕ be a *u-shaped function*, i.e. a continuous function $\phi : \mathbb{R} \to [1, \infty)$ that is nonincreasing on $(-\infty, 0)$ and nondecreasing on $(0, \infty)$. Then

$$d_{(\phi)}(\mu, \nu) := \sup_{y \in \mathbb{R}} |F_\mu(y) - F_\nu(y)| \, \phi(y) \tag{4}$$

defines a metric on the set $\mathcal{M}_1^{(\phi)}(\mathbb{R})$ of all probability measures μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ for which $d_{(\phi)}(\mu, \delta_0) < \infty$. We will refer to $d_{(\phi)}$ as *Kolmogorov ϕ-metric*. Notice that $\mu \in \mathcal{M}_1^{(\phi)}(\mathbb{R})$ if $\int \phi \, d\mu < \infty$, and that $d_{(\phi)}$ is just the classical Kolmogorov metric for $\phi := \mathbb{1}$.

In [2, 38] it is demonstrated that many L- and V-functionals T as well as many coherent risk functionals T are continuous w.r.t. $d_{(\phi)}$ with ϕ depending on the particular functional T; see also Example 3.4 for some simple examples. So, in view of the Hampel-type Theorem 2.4, for a given class $\mathcal{P}$ of probability measures on $(\Omega, \mathcal{F})$, the corresponding sequence $(\widehat{T}_n)$ of plug-in estimators is qualitatively $\mathcal{P}$-marginally robust w.r.t. $(d_{(\phi)}, d'_{\text{Proh}})$ at any $\mathbb{P} \in \mathcal{P}$ if $\mathcal{P}$ admits the UGC property w.r.t. $d_{(\phi)}$ in the sense of Definition 2.3. The following Theorem 3.1 shows that the latter is true if under every $\mathbb{P} \in \mathcal{P}$ the coordinates of the coordinate process (X_i) are identically distributed and

strongly mixing with uniformly (in $\mathbb{P} \in \mathcal{P}$) decaying mixing coefficients $(\alpha_\mathbb{P}(n))$ and if the class of marginal distributions $\mathcal{P}_1$ is uniformly ϕ-integrating. It is remarkable that the common rate of decay of the mixing coefficients may be arbitrarily slow.

Theorem 3.1 (UGC property w.r.t. $d_{(\phi)}$) *Let ϕ be a u-shaped function, and $\mathcal{P}$ be a class of probability measures on $(\Omega, \mathcal{F})$ such that*

(a) $\mathbb{P}_1 = \mathbb{P}_2 = \cdots$ *for all* $\mathbb{P} \in \mathcal{P}$,

(b) $\lim_{K \to \infty} \sup_{\mathbb{P} \in \mathcal{P}} \int \phi(y) \mathbb{1}_{\phi(y) \geq K} \, \mathbb{P}_1(dy) = 0$,

(c) $\lim_{n \to \infty} \sup_{\mathbb{P} \in \mathcal{P}} \alpha_\mathbb{P}(n) = 0$.

Then, for every $\delta > 0$,

$$\lim_{n \to \infty} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\, d_{(\phi)}(\widehat{m}_n, \mathbb{P}_1) \geq \delta \,] = 0. \tag{5}$$

The proof of Theorem 3.1 can be found in Section 5. Notice that (c) implies in particular that the coordinate process (X_i) is strongly mixing under every $\mathbb{P} \in \mathcal{P}$.

Remark 3.2 (i) Condition (b) is always fulfilled if ϕ is bounded, in particular if $d_{(\phi)}$ is the classical Kolmogorov metric ($\phi = \mathbb{1}$).

(ii) Condition (b) is also fulfilled if one can find some function $w : \mathbb{R} \to [0, \infty)$ such that $\lim_{|x| \to \infty} w(x)/\phi(x) = \infty$ and $\sup_{\mathbb{P} \in \mathcal{P}} \int w(y) \, \mathbb{P}_1(dy) < \infty$; cf. [22, Lemma 4.1]. $\diamond$

Remark 3.3 Since the Lévy metric $d_{\text{Lévy}}$ (defined in (6) below) generates the weak topology (cf. [19, p. 25]) and the Kolmogorov metric $d_{(\mathbb{1})}$ dominates the Lévy metric, i.e. $d_{\text{Lévy}} \leq d_{(\mathbb{1})}$ (cf. [19, p. 34]), we obtain that every functional T which is weakly continuous at some $\mu \in \mathcal{M}_1(\mathbb{R})$ is also continuous w.r.t. $d_{(\mathbb{1})}$ at μ. Further notice that qualitative robustness w.r.t. $d_{\text{Lévy}}$ implies qualitative robustness w.r.t. $d_{(\mathbb{1})}$. On the other hand, the UGC property w.r.t. $d_{(\mathbb{1})}$ implies the UGC property w.r.t. $d_{\text{Lévy}}$. $\diamond$

Example 3.4 (i) It is well-known that for fixed $\alpha \in (0, 1)$ the lower α-quantile functional $T_\alpha(\mu) := F_\mu^\leftarrow(\alpha)$ is continuous w.r.t. the weak topology at $\mu \in \mathcal{M}_1(\mathbb{R})$ when $F_\mu^\leftarrow$ is continuous at α; see, for instance, [33, Lemma 21.2]. According to the first part of Remark 3.3, T_α is in particular continuous w.r.t. the Kolmogorov metric $d_{(\mathbb{1})}$ at this μ.

(ii) The mean functional $T^{(1)}(\mu) := \int y \, \mu(dy)$ is continuous on $\mathcal{M}^{(\phi)}$ w.r.t. $d_{(\phi)}$ for any ϕ satisfying $\int 1/\phi(y) \, dy < \infty$. This follows from the inequality $|T^{(1)}(\mu) - T^{(1)}(\nu)| \leq \int |F_\mu(y) - F_\nu(y)| \, dy$ which holds for all $\mu, \nu \in \mathcal{M}^{(\phi)}$. For instance, we may choose $\phi(y) = (1 + |y|)^{1+\varepsilon}$ for arbitrarily small $\varepsilon > 0$. In this case, condition (b) in Theorem 3.1 holds when $\sup_{\mathbb{P} \in \mathcal{P}} \int |y|^{1+\varepsilon'} \, \mathbb{P}_1(dy) < \infty$ for some $\varepsilon' > \varepsilon$; cf. Remark 3.2 (ii).

(iii) The second moment functional $T^{(2)}(\mu) := \int y^2 \, \mu(dy)$, and thus the variance functional, is continuous on $\mathcal{M}^{(\phi)}$ w.r.t. $d_{(\phi)}$ for any ϕ satisfying $\int |y|/\phi(y) \, dy < \infty$. This follows from the inequality $|T^{(2)}(\mu) - T^{(2)}(\nu)| \leq 2 \int |F_\mu(y) - F_\nu(y)| \, |y| \, dy$ which holds for all $\mu, \nu \in \mathcal{M}^{(\phi)}$. For instance, we may choose $\phi(y) = (1+|y|)^{2+\varepsilon}$ for arbitrarily small $\varepsilon > 0$. In this case, condition (b) in Theorem 3.1 holds when $\sup_{\mathbb{P} \in \mathcal{P}} \int |y|^{2+\varepsilon'} \, \mathbb{P}_1(dy) < \infty$ for some $\varepsilon' > \varepsilon$; cf. Remark 3.2 (ii). $\diamond$

Example 3.5 *(Linear processes)* Let $(Z_s)_{s \in \mathbb{Z}}$ be a sequence of i.i.d. random variables on any probability space, assume that $|Z_1|$ has a finite expectation denoted by L, and assume that the distribution of Z_1 admits a Lebesgue density f for which $\int |f(y+h) - f(y)| \, dy < M|h|$ for all $h \in \mathbb{R}$ and some constant $M > 0$. Define the linear process $X_t^a := \sum_{s=0}^\infty a_s Z_{t-s}$, $t \in \mathbb{N}$, for any real sequence $a = (a_s)_{s \in \mathbb{N}_0}$, and let A be the class of all real sequences a for which $a_0 = 1$ and $\sum_{s=0}^\infty a_s Z_{t-s}$ is almost surely absolutely convergent for every $t \in \mathbb{N}$. Results in [26] imply that, if $a \in A$ satisfies $\sum_{s=0}^\infty a_s z^s \neq 0$ for all z with $|z| \leq 1$, and $\sum_{u=1}^\infty \sum_{s=u}^\infty |a_s| < \infty$, then (X_t^a) is strongly mixing with mixing coefficients $(\alpha_a(n))$ satisfying $\alpha_a(n) \leq (2ML \sum_{s=0}^\infty |b_s(a)|) \sum_{u=n}^\infty \sum_{s=u}^\infty |a_s|$, where $b_s(a)$ is the coefficient of z^s in the power series expansion of $z \mapsto 1/\sum_{s=0}^\infty a_s z^s$, and $\sum_{s=0}^\infty |b_s(a)| < \infty$; cf. the Appendix B. If we denote by $\mathbb{P}^a$ the law of (X_t^a) on $\mathbb{R}^\mathbb{N}$, then, of course, the coordinate process (X_t) on $\mathbb{R}^\mathbb{N}$ is also strongly mixing under $\mathbb{P}^a$ with mixing coefficients $(\alpha_{\mathbb{P}^a}(n))$ satisfying $\alpha_{\mathbb{P}^a}(n) \leq (2ML \sum_{s=0}^\infty |b_s(a)|) \sum_{u=n}^\infty \sum_{s=u}^\infty |a_s|$.

Now, let A' be any subset of A such that

(i) $\sum_{s=0}^\infty a_s z^s \neq 0$ for all z with $|z| \leq 1$, for every $a \in A'$,

(ii) $\lim_{n \to \infty} \sup_{a \in A'} \sum_{u=n}^\infty \sum_{s=u}^\infty |a_s| = 0$,

(iii) $\sup_{a \in A'} \sum_{s=0}^\infty |b_s(a)| < \infty$.

Then we obtain from the statement above that the class $\mathcal{P} = \mathcal{P}' := \{\mathbb{P}^a : a \in A'\}$ satisfies condition (c) in Theorem 3.1. Moreover, condition (a) in Theorem 3.1 is fulfilled for $\mathcal{P} = \mathcal{P}'$ anyway, and condition (b) in Theorem 3.1 is always fulfilled when $\phi = \mathbb{1}$. Therefore, Theorem 3.1 shows that $\mathcal{P}'$ admits the UGC property for the Kolmogorov metric $d_{(\mathbb{1})}$. In particular, the Hampel-type Theorem 2.4 implies that any $d_{(\mathbb{1})}$-continuous functional T is qualitatively $\mathcal{P}'$-marginally robust at any $\mathbb{P}^a \in \mathcal{P}'$ w.r.t. $(d_{(\mathbb{1})}, d'_{\mathrm{Proh}})$. That is, for every $\varepsilon > 0$ there are some $\delta > 0$ and $n_0 \in \mathbb{N}$ such that

$$\mathbb{P}^{a'} \in \mathcal{P}', \quad d_{(\mathbb{1})}(\mathbb{P}_1^a, \mathbb{P}_1^{a'}) \leq \delta \quad \Longrightarrow \quad d'_{\mathrm{Proh}}(\mathbb{P}^a \circ \widehat{T}_n^{-1}, \mathbb{P}^{a'} \circ \widehat{T}_n^{-1}) \leq \varepsilon \quad \forall \, n \geq n_0.$$

To get a feeling for the condition on the left, it is appealing to find preferably sharp conditions on $a' = (a_s')_{s \in \mathbb{N}_0}$ and the distribution of Z_0 under which the distance $d_{(\mathbb{1})}(\mathbb{P}_1^a, \mathbb{P}_1^{a'})$ does not exceed a given $\delta > 0$. This is an interesting problem on its own, and we will not enlarge upon this here. We will only mention that, if $\sum_{s=0}^\infty d_{(\mathbb{1})}(\mu_{a_s}, \mu_{a_s'}) < \infty$ with μ_{a_s}

and $\mu_{a'_s}$ the laws of $a_s Z_0$ and $a'_s Z_0$, respectively, then, using the convolution formula, one easily obtains the estimate $d_{(\mathbb{1})}(\mathbb{P}_1^a, \mathbb{P}_1^{a'}) \le \sum_{s=0}^{\infty} d_{(\mathbb{1})}(\mu_{a_s}, \mu_{a'_s})$ from where on can derive some respective conditions. However, there might be more sophisticated approaches. $\diamond$

Example 3.6 *(ARMA(1,1) processes)* To illustrate conditions (i)–(iii) in Example 3.5, let us consider for any real ϕ_1, θ_1 an ARMA(1,1) process $X_t^{\phi_1, \theta_1} = \phi_1 X_{t-1}^{\phi_1, \theta_1} + Z_t + \theta_1 Z_{t-1}$ based on a given sequence $(Z_s)_{s \in \mathbb{Z}}$ of square-integrable and centered i.i.d. random variables. Moreover, let $0 < c < 1$ be arbitrary but fixed. It is discussed in detail in Example B.4 in the Appendix B that if $|\phi_1|, |\theta_1| \le c$ and $\phi_1 \ne -\theta_1$, then the ARMA process $(X_t^{\phi_1, \theta_1})$ can be represented as a linear process $\sum_{s=0}^{\infty} a_s(\phi_1, \theta_1) Z_{t-s}$ with $a_0(\phi_1, \theta_1) = 1$. Letting $A'_c := \{(a_s(\phi_1, \theta_1))_{s \in \mathbb{N}_0} : |\phi_1|, |\theta_1| \le c$ and $\phi_1 \ne -\theta_1\}$, Example B.4 also shows that

(i) $\sum_{s=0}^{\infty} a_s z^s \ne 0$ for all z with $|z| \le 1$, for every $a \in A'_c$,

(ii) $\sup_{a \in A'_c} \sum_{u=n}^{\infty} \sum_{s=u}^{\infty} |a_s| \le (2/(1-c)^2) c^n$ for all $n \in \mathbb{N}$,

(iii) $\sup_{a \in A'_c} \sum_{s=0}^{\infty} |b_s(a)| \le 1 + 2c/(1-c)$.

Thus, denoting by $\mathbb{P}^{\phi_1, \theta_1}$ the law of the ARMA(1,1) process $(X_t^{\phi_1, \theta_1})$ based on the given noise (Z_s) and with coefficients ϕ_1, θ_1, the discussion in Example 3.5 shows that the class $\mathcal{P} = \mathcal{P}_c := \{\mathbb{P}^{\phi_1, \theta_1} : |\phi_1|, |\theta_1| \le c$ and $\phi_1 \ne -\theta_1\}$ satisfies condition (c) in Theorem 3.1, because $\mathcal{P}_c$ is nothing but the class of laws $\mathbb{P}^a$ on $\mathbb{R}^{\mathbb{N}}$ of all linear processes $\sum_{s=0}^{\infty} a_s Z_{t-s}$ with $a \in A'_c$. $\diamond$

Recall from [19, p. 25] that the Lévy metric

$$d_{\text{Lévy}}(\mu, \nu) := \inf\{\varepsilon > 0 : F_\mu(x - \varepsilon) - \varepsilon \le F_\nu(x) \le F_\mu(x + \varepsilon) + \varepsilon \text{ for all } x \in \mathbb{R}\} \quad (6)$$

(with F_μ, F_ν the distribution functions of $\mu, \nu \in \mathcal{M}_1(\mathbb{R})$) generates the weak topology on $\mathcal{M}_1(\mathbb{R})$. Since the classical Kolmogorov metric $d_{(\mathbb{1})}$ dominates the Lévy metric (cf. [19, p. 34]), we immediately obtain the following corollary to Theorem 3.1 with $\phi := \mathbb{1}$, taking Remark 3.2 (i) into account.

Corollary 3.7 (UGC property w.r.t. $d_{\text{Lévy}}$) *Let $\mathcal{P}$ be a class of probability measures on $(\Omega, \mathcal{F})$ satisfying conditions (a) and (c) of Theorem 3.1. Then we have (5) with $d_{(\phi)}$ replaced by $d_{\text{Lévy}}$.*

Obviously, Example 3.5 remains valid (except the estimate of $d_{(\mathbb{1})}(\mathbb{P}_1^a, \mathbb{P}_1^{a'})$ at the end) when replacing the Kolmogorov metric $d_{(\mathbb{1})}$ by the Lévy metric $d_{\text{Lévy}}$.

3.2. Strong mixing, and ψ-weak topology

Let E be a locally compact and second countable Hausdorff space (in this case E is in particular a Polish space), $\mathcal{E}$ be the corresponding Borel σ-field, and $\psi : E \to [0, \infty)$ be a continuous function satisfying $\psi \geq 1$ outside some compact set. Let $\mathcal{M}_1^\psi(E)$ be the set of all probability measures μ on $(E, \mathcal{E})$ satisfying $\int \psi \, d\mu < \infty$, and $C_\psi(E)$ be the space of all continuous functions on E for which $\|f/(1+\psi)\|_\infty < \infty$, where $\|g\|_\infty := \sup_{y \in E} |g(y)|$ for any function $g : E \to \mathbb{R}$. The ψ-*weak topology* on $\mathcal{M}_1^\psi(E)$ is the coarsest topology for which the mappings $\mu \mapsto \int f \, d\mu$, $f \in C_\psi(E)$, are continuous; cf. Section A.6 in [14]. Clearly, the ψ-weak topology is finer than the weak topology, and the two topologies coincide if and only if ψ is bounded. It follows from Lemma 3.4 (i)$\Leftrightarrow$(iii) in [21] (which still holds when replacing $\mathbb{R}^d$ by some Polish space) that the metric

$$d_{\psi,\mathrm{vag}}(\mu,\nu) := d_{\mathrm{vag}}(\mu,\nu) + \Big| \int \psi \, d\mu - \int \psi \, d\nu \Big|, \qquad \mu, \nu \in \mathcal{M}_1^\psi(E) \tag{7}$$

metrizes the ψ-weak topology when d_{vag} metrizes the vague topology on $\mathcal{M}_1^\psi(E)$. For d_{vag} we may and do choose

$$d_{\mathrm{vag}}(\mu,\nu) := \sum_{k=1}^\infty \frac{1}{2^k} \Big\{ 1 \wedge \Big| \int f_k \, d\mu - \int f_k \, d\nu \Big| \Big\}, \qquad \mu, \nu \in \mathcal{M}_1^\psi(E) \tag{8}$$

for some countable and $\|\cdot\|_\infty$-dense subset $\{f_k\}_{k \in \mathbb{N}}$ of the space $C_{\mathsf{c}}(E)$ of all continuous functions on E with compact support; cf. the proof of Theorem 31.5 in [1].

Remark 3.8 Any locally compact and second countable Hausdorff space E is σ-compact (cf. Example 2 in Section 29 of [1]), i.e. there exists a sequence (K_n) of compact subsets of E such that $K_n \uparrow E$ and every compact set K is contained in finally all K_n. So by Urysohn's lemma one can find functions $e_n \in C_{\mathsf{c}}(E)$, $n \in \mathbb{N}$, such that $0 \leq e_n \leq 1$ and $e_n = 1$ on K_n for all $n \in \mathbb{N}$. If $\{\widetilde{f_l}\}_{l \in \mathbb{N}}$ denotes any countable and $\|\cdot\|_\infty$-dense subset of $C_{\mathsf{c}}(E)$, then the set $\{f_k\}_{k \in \mathbb{N}}$ in (8) can be chosen as

$$\{f_k\}_{k \in \mathbb{N}} := \{\widetilde{f_l}\}_{l \in \mathbb{N}} \cup \{\widetilde{f_l} \, e_n\}_{l,n \in \mathbb{N}} \cup \{e_n\}_{n \in \mathbb{N}}.$$

This gets clear from the elaborations in the proof of Theorem 31.5 in [1]. $\qquad\qquad \diamond$

The co-variance functional $T(\mu) := \int_{\mathbb{R}^2} (x_1 - \int_{\mathbb{R}} x \, \mu_1(dx))(x_2 - \int_{\mathbb{R}} x \, \mu_2(dx)) \, \mu(d(x_1, x_2))$ (μ_1 and μ_2 denote the marginal distributions of μ), for instance, is clearly ψ-weakly continuous for $\psi(x) := |x|^2$ and $E = \mathbb{R}^2$. So, in view of the Hampel-type Theorem 2.4, for a given class $\mathcal{P}$ of probability measures on $(\Omega, \mathcal{F})$, the corresponding sequence $(\widehat{T}_n)$ of plug-in estimators, i.e. the sequence of the sample co-variances, is qualitatively $\mathcal{P}$-marginally robust w.r.t. $(d_{\psi,\mathrm{vag}}, d'_{\mathrm{Proh}})$ at any $\mathbb{P} \in \mathcal{P}$ if $\mathcal{P}$ admits the UGC property w.r.t. $d_{\psi,\mathrm{vag}}$ in the sense of Definition 2.3. The following Theorem 3.9 shows that the latter is true under similar conditions as imposed in Theorem 3.1.

Theorem 3.9 (UGC property w.r.t. $d_{\psi,\mathrm{vag}}$) *Let E be a locally compact and second countable Hausdorff space, $\mathcal{E}$ be the corresponding Borel σ-field, and $\psi : E \to [1, \infty)$ be a continuous function. Let $\mathcal{P}$ be a class of probability measures on $(\Omega, \mathcal{F})$ such that*

(a) $\mathbb{P}_1 = \mathbb{P}_2 = \cdots$ *for all* $\mathbb{P} \in \mathcal{P}$,

(b) $\lim_{K \to \infty} \sup_{\mathbb{P} \in \mathcal{P}} \int \psi(y) \mathbb{1}_{\psi(y) \geq K} \, \mathbb{P}_1(dy) = 0$,

(c) $\lim_{n \to \infty} \sup_{\mathbb{P} \in \mathcal{P}} \alpha_{\mathbb{P}}(n) = 0$.

Then, for every $\delta > 0$,

$$\lim_{n \to \infty} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\, d_{\psi,\mathrm{vag}}(\widehat{m}_n, \mathbb{P}_1) \geq \delta \,] \; = \; 0. \tag{9}$$

The proof of Theorem 3.1 can be found in Section 6. Notice that (c) implies in particular that the coordinate process (X_i) is strongly mixing under every $\mathbb{P} \in \mathcal{P}$.

Remark 3.10 (i) Notice that condition (b) is always fulfilled if ψ is bounded, i.e. if $d_{\psi,\mathrm{vag}}$ metrizes the classical weak topology.

(ii) In the case $E = \mathbb{R}$, condition (b) is fulfilled if there is some $w : \mathbb{R} \to [0, \infty)$ such that $\lim_{|x| \to \infty} w(x)/\phi(x) = \infty$ and $\sup_{\mathbb{P} \in \mathcal{P}} \int w(y) \, \mathbb{P}_1(dy) < \infty$; cf. [22, Lemma 4.1]. $\quad\diamond$

Recall from (6) the definition of the Lévy metric $d_{\mathrm{Lévy}}$. Since $d_{\mathrm{Lévy}}$ generates the weak topology on $\mathcal{M}_1(\mathbb{R})$, the metric

$$d_{\psi,\mathrm{Lévy}}(\mu, \nu) \; := \; d_{\mathrm{Lévy}}(\mu, \nu) + \left| \int \psi \, d\mu - \int \psi \, d\nu \right|, \qquad \mu, \nu \in \mathcal{M}_1^{\psi}(\mathbb{R}) \tag{10}$$

generates the ψ-weak topology on $\mathcal{M}_1^{\psi}(\mathbb{R})$. The following corollary is an immediate consequence of Theorem 3.9 (in fact of (25) in its proof) and Corollary 3.7.

Corollary 3.11 (UGC property w.r.t. $d_{\psi,\mathrm{Lévy}}$) *Let $E = \mathbb{R}$, and $\mathcal{P}$ be a class of probability measures on $(\Omega, \mathcal{F})$ satisfying conditions (a)–(c) of Theorem 3.9. Then we have (9) with $d_{\psi,\mathrm{vag}}$ replaced by $d_{\psi,\mathrm{Lévy}}$.*

Example 3.12 (i) It was already mentioned in Example 3.4 (i) that for fixed $\alpha \in (0, 1)$ the lower α-quantile functional is continuous at $\mu \in \mathcal{M}_1(\mathbb{R})$ w.r.t. the classical weak topology, i.e. w.r.t. the ψ-weak topology with $\psi = \mathbb{1}$, provided $F_{\mu}^{\leftarrow}$ is continuous at α.

(ii) It is easily seen that the mean functional and the variance functional are continuous w.r.t. the ψ-weak topology on $\mathcal{M}_1^{\psi}(\mathbb{R})$ for $\psi = |x|$ and $\psi = |x|^2$, respectively.

(iii) It is demonstrated in [22] that the statistical functional $T(\mu) = \rho(X_{\mu})$ corresponding to any law-invariant convex risk measure ρ defined on an Orlicz space with continuous Young function Ψ satisfying the Δ_2-condition (meaning that there are $C, x_0 > 0$ such that $\Psi(2x) \leq C\Psi(x)$ for all $x \geq x_0$) is continuous w.r.t. the ψ-weak topology on $\mathcal{M}_1^{\psi}(\mathbb{R})$ for $\psi(\cdot) = \Psi(|\cdot|)$. $\quad\diamond$

4. Proof of Theorem 2.4

We adapt the proof of the classical Hampel theorem as given in [19, 20]. For the reader's convenience, we first of all recall Strassen's theorem (as formulated in Theorem 2.4.7 in [19]; the proof is contained in the seminal paper [32]) whose implication (ii)⇒(i) is the key for the proof of Theorem 2.4.

Theorem 4.1 (Strassen) *Let $\mathbf{T}$ be a Polish space equipped with the corresponding Borel σ-field $\mathcal{T}$, and $d_{\mathbf{T}}$ be any complete and separable metric generating the topology on $\mathbf{T}$. Then, for any two probability measures μ_1, μ_2 on $(\mathbf{T}, \mathcal{T})$ and any $\varepsilon, \delta > 0$, the following two statements are equivalent:*

(i) *For every $A \in \mathcal{T}$ we have*
$$\mu_1[A] \le \mu_2[A^\delta] + \varepsilon,$$
where $A^\delta := \{t \in \mathbf{T} : \inf_{a \in A} d_{\mathbf{T}}(t, a) \le \delta\}$.

(ii) *There is some probability measure μ on $(\mathbf{T} \times \mathbf{T}, \mathcal{T} \times \mathcal{T})$ such that $\mu \circ \pi_1^{-1} = \mu_1$, $\mu \circ \pi_2^{-1} = \mu_2$, and*
$$\mu\Big[\Big\{(t_1, t_2) \in \mathbf{T} \times \mathbf{T} : d_{\mathbf{T}}(t_1, t_2) \le \delta\Big\}\Big] \ge 1 - \varepsilon,$$
where $\pi_i : \mathbf{T} \times \mathbf{T} \to \mathbf{T}$ denotes the projection on the ith coordinate, $i = 1, 2$.

To prove Theorem 2.4, we have to show that for every $\varepsilon > 0$ there are some $\delta > 0$ and $n_0 \in \mathbb{N}$ such that
$$\mathbb{Q} \in \mathcal{P}, \ d(\mathbb{P}_1, \mathbb{Q}_1) \le \delta \quad \Longrightarrow \quad d'_{\text{Proh}}(\mathbb{P} \circ \widehat{T}_n^{-1}, \mathbb{Q} \circ \widehat{T}_n^{-1}) \le \varepsilon \quad \forall\, n \ge n_0. \qquad (11)$$

Since
$$d'_{\text{Proh}}(\mathbb{P} \circ \widehat{T}_n^{-1}, \mathbb{Q} \circ \widehat{T}_n^{-1}) \le d'_{\text{Proh}}(\mathbb{P} \circ \widehat{T}_n^{-1}, \delta_{T(\mathbb{P}_1)}) + d'_{\text{Proh}}(\delta_{T(\mathbb{P}_1)}, \mathbb{Q} \circ \widehat{T}_n^{-1})$$

(with $\delta_{T(\mathbb{P}_1)}$ the dirac measure on $(\mathbf{T}, \mathcal{T})$ with atom $T(\mathbb{P}_1)$), for (11) it suffices to show that for every $\varepsilon > 0$ there are some $\delta > 0$ and $n_0 \in \mathbb{N}$ such that
$$\mathbb{Q} \in \mathcal{P}, \ d(\mathbb{P}_1, \mathbb{Q}_1) \le \delta \quad \Longrightarrow \quad d'_{\text{Proh}}(\delta_{T(\mathbb{P}_1)}, \mathbb{Q} \circ \widehat{T}_n^{-1}) \le \varepsilon/2 \quad \forall\, n \ge n_0. \qquad (12)$$

The remainder of the proof is divided into two steps. In Step 1, we will verify that for (12) it suffices to show that for every $\varepsilon > 0$ there are some $\delta > 0$ and $n_0 \in \mathbb{N}$ such that
$$\mathbb{Q} \in \mathcal{P}, \ d(\mathbb{P}_1, \mathbb{Q}_1) \le \delta \quad \Longrightarrow \quad \mathbb{Q}\Big[\Big\{x \in \Omega : d_{\mathbf{T}}(T(\mathbb{P}_1), \widehat{T}_n(x)) \le \frac{\varepsilon}{2}\Big\}\Big] \ge 1 - \frac{\varepsilon}{2} \quad \forall\, n \ge n_0. \tag{13}$$

In Step 2, we will verify (13).

Step 1. We note that the right-hand side in (13) is equivalent to

$$\left(\delta_{\mathbf{T}(\mathbb{P}_1)} \times (\mathbb{Q} \circ \widehat{T}_n^{-1})\right)\left[\left\{(t_1, t_2) \in \mathbf{T} \times \mathbf{T} : d_{\mathbf{T}}(t_1, t_2) \leq \frac{\varepsilon}{2}\right\}\right] \geq 1 - \frac{\varepsilon}{2} \quad \forall n \geq n_0. \quad (14)$$

In view of the implication (ii)⇒(i) in Strassen's Theorem 4.1 (with $\mu := \delta_{T(\mathbb{P}_1)} \times (\mathbb{Q} \circ \widehat{T}_n^{-1})$ and $\widetilde{\varepsilon} := \widetilde{\delta} := \varepsilon/2$), condition (14) implies

$$\delta_{T(\mathbb{P}_1)}[A] \leq \mathbb{Q} \circ \widehat{T}_n^{-1}[A^{\varepsilon/2}] + \varepsilon/2 \qquad \forall A \in \mathcal{T}, n \geq n_0,$$

i.e.

$$d'_{\mathrm{Proh}}(\delta_{T(\mathbb{P}_1)}, \mathbb{Q} \circ \widehat{T}_n^{-1}) \leq \varepsilon/2 \qquad \forall n \geq n_0.$$

That is, the right-hand side in (13) implies the right-hand side in (12).

Step 2. To verify (13), we pick $\varepsilon > 0$. Since T is $(d, d_{\mathbf{T}})$-continuous at $\mathbb{P}_1$, we can find some $\delta > 0$ such that for every $x \in \Omega$ and $n \in \mathbb{N}$

$$d(\mathbb{P}_1, \widehat{m}_n(x)) \leq 2\delta \quad \Longrightarrow \quad d_{\mathbf{T}}(T(\mathbb{P}_1), T(\widehat{m}_n(x))) \leq \varepsilon/2; \quad (15)$$

recall that the class of all empirical probability measures $\widehat{m}_n(x)$ was assumed to be contained in $\mathcal{M}$. Now, for any $\mathbb{Q} \in \mathcal{P}$ satisfying $d(\mathbb{P}_1, \mathbb{Q}_1) \leq \delta$, we obtain with the help of the UGC property of $\mathcal{P}$ w.r.t. d, of the implication

$$d(\mathbb{Q}_1, \widehat{m}_n(x)) \leq \delta \quad \Longrightarrow \quad d(\mathbb{P}_1, \widehat{m}_n(x)) \; [\leq d(\mathbb{P}_1, \mathbb{Q}_1) + d(\mathbb{Q}_1, \widehat{m}_n(x))] \; \leq \; d(\mathbb{P}_1, \mathbb{Q}_1) + \delta$$

and of (15) that

$$
\begin{aligned}
1 - \frac{\varepsilon}{2} \; &\leq \; \mathbb{Q}\left[\left\{x \in \Omega : d(\mathbb{Q}_1, \widehat{m}_n(x)) \leq \delta\right\}\right] \\
&\leq \; \mathbb{Q}\left[\left\{x \in \Omega : d(\mathbb{P}_1, \widehat{m}_n(x)) \leq d(\mathbb{P}_1, \mathbb{Q}_1) + \delta\right\}\right] \\
&\leq \; \mathbb{Q}\left[\left\{x \in \Omega : d(\mathbb{P}_1, \widehat{m}_n(x)) \leq 2\delta\right\}\right] \\
&\leq \; \mathbb{Q}\left[\left\{x \in \Omega : d_{\mathbf{T}}(T(\mathbb{P}_1), T(\widehat{m}_n(x))) \leq \varepsilon/2\right\}\right] \\
&= \; \mathbb{Q}\left[\left\{x \in \Omega : d_{\mathbf{T}}(T(\mathbb{P}_1), \widehat{T}_n(x)) \leq \varepsilon/2\right\}\right]
\end{aligned}
$$

for all $n \geq n_0$ and some $n_0 = n_0(\varepsilon) \in \mathbb{N}$, where n_0 can be chosen independently of $\mathbb{Q}$ by the UGC property of $\mathcal{P}$. Thus, (13) holds. This completes the proof of Theorem 2.4.

5. Proof of Theorem 3.1

We will only show that for every $\delta > 0$

$$\lim_{n \to \infty} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left[\sup_{y \in (-\infty, 0]} |\widehat{F}_n(y) - F_{\mathbb{P}}(y)| \, \phi(y) \geq \delta\right] = 0, \quad (16)$$

where $\widehat{F}_n$ and $F_{\mathbb{P}}$ denote the empirical distribution function of $X_1, \ldots, X_n$ and the distribution function of the marginal distribution $\mathbb{P}_1$, respectively. The analogous result for the positive real line can be shown in the same way. We will proceed in five steps. For every $\mathbb{P} \in \mathcal{P}$, let the functions $w_{\mathbb{P}} : [0, 1] \to [0, \infty]$ and $h_{\mathbb{P}} : [0, 1] \to [0, \infty)$ be defined by

$$
\begin{aligned}
w_{\mathbb{P}}(t) &:= \phi(F_{\mathbb{P}}^{\leftarrow}(t)) \mathbb{1}_{[0, F_{\mathbb{P}}(0)]}(t), \\
h_{\mathbb{P}}(t) &:= \int_0^t w_{\mathbb{P}}(s) \, ds.
\end{aligned}
$$

Of course, the functions $h_{\mathbb{P}}$ are continuous, nondecreasing and satisfy $h_{\mathbb{P}}(0) = 0$ for all $\mathbb{P} \in \mathcal{P}$. By assumption (b), we also have $\sup_{\mathbb{P} \in \mathcal{P}} h_{\mathbb{P}}(1) \, (\leq \sup_{\mathbb{P} \in \mathcal{P}} \int \phi \, d\mathbb{P}_1) < \infty$.

Step 1. As the first step, we will show that the family of functions $\{h_{\mathbb{P}}\}_{\mathbb{P} \in \mathcal{P}}$ is uniformly equicontinuous on $[0, 1]$. Denote by $J_{\mathbb{P}}$ the set of all $y \in (-\infty, 0]$ at which $F_{\mathbb{P}}$ has a jump, and by $I_{\mathbb{P}} \subset [0, 1]$ the range of $F_{\mathbb{P}}$. Noting that $F_{\mathbb{P}}(F_{\mathbb{P}}^{\leftarrow}(s)) = s$ if and only if $s \in I_{\mathbb{P}}$ and that $F_{\mathbb{P}}^{\leftarrow}(s) \in J_{\mathbb{P}}$ for all $s \in [0, 1] \setminus I_{\mathbb{P}}$ (cf. [30, p. 113]), we obtain for any $K > 0$

$$
\begin{aligned}
&\sup_{\mathbb{P} \in \mathcal{P}} h_{\mathbb{P}}(t) \\
&= \sup_{\mathbb{P} \in \mathcal{P}} \int \phi(F_{\mathbb{P}}^{\leftarrow}(s)) \mathbb{1}_{[0, F_{\mathbb{P}}(0) \wedge t] \cap I_{\mathbb{P}}}(s) \, ds \\
&\quad + \sup_{\mathbb{P} \in \mathcal{P}} \int \phi(F_{\mathbb{P}}^{\leftarrow}(s)) \mathbb{1}_{[0, F_{\mathbb{P}}(0) \wedge t] \setminus I_{\mathbb{P}}}(s) \, ds \\
&\leq \sup_{\mathbb{P} \in \mathcal{P}} \int \phi(F_{\mathbb{P}}^{\leftarrow}(s)) \mathbb{1}_{[0, F_{\mathbb{P}}(0) \wedge t] \cap I_{\mathbb{P}}}(F_{\mathbb{P}}(F_{\mathbb{P}}^{\leftarrow}(s))) \, ds \\
&\quad + \sup_{\mathbb{P} \in \mathcal{P}} \sum_{y \in J_{\mathbb{P}}} \phi(y) \Big((F_{\mathbb{P}}(y) \wedge t) - (F_{\mathbb{P}}(y-) \wedge t) \Big) \\
&\leq \sup_{\mathbb{P} \in \mathcal{P}} \int_{\mathbb{R}_-} \phi(y) \, \mathbb{1}_{\phi(y) \geq K} \, \mathbb{P}_1(dy) + \sup_{\mathbb{P} \in \mathcal{P}} K \int_{\mathbb{R}_-} \mathbb{1}_{[0, t]}(F_{\mathbb{P}}(y)) \, \mathbb{P}_1(dy) \\
&\quad + \sup_{\mathbb{P} \in \mathcal{P}} \int_{\mathbb{R}_-} \phi(y) \, \mathbb{1}_{\phi(y) \geq K} \, \mathbb{P}_1(dy) + \sup_{\mathbb{P} \in \mathcal{P}} K \sum_{y \in J_{\mathbb{P}}} \Big((F_{\mathbb{P}}(y) \wedge t) - (F_{\mathbb{P}}(y-) \wedge t) \Big) \\
&\leq \sup_{\mathbb{P} \in \mathcal{P}} \int \phi(y) \, \mathbb{1}_{\phi(y) \geq K} \, \mathbb{P}_1(dy) + \sup_{\mathbb{P} \in \mathcal{P}} K \, \mathbb{P}_1[\{z : F_{\mathbb{P}}(z) \leq t\}] \\
&\quad + \sup_{\mathbb{P} \in \mathcal{P}} \int \phi(y) \, \mathbb{1}_{\phi(y) \geq K} \, \mathbb{P}_1(dy) + K \, t \\
&= 2 \sup_{\mathbb{P} \in \mathcal{P}} \int \phi(y) \, \mathbb{1}_{\phi(y) \geq K} \, \mathbb{P}_1(dy) + 2K \, t.
\end{aligned}
$$

Now, by assumption (b) we may choose $K = K_\varepsilon > 0$ such that the first summand is bounded above by $\varepsilon/2$. In particular, $h_{\mathbb{P}}(t) \leq \varepsilon$ for all $t \leq \varepsilon/(4K_\varepsilon)$ and $\mathbb{P} \in \mathcal{P}$. That is, $h_{\mathbb{P}}$ is (right) continuous at 0 uniformly in $\mathbb{P} \in \mathcal{P}$. The uniform (right) continuity at 0 moreover implies uniform equicontinuity of the family $\{h_{\mathbb{P}}\}_{\mathbb{P} \in \mathcal{P}}$ on $[0, 1]$, because for

every $t \in [0,1]$ and $\Delta \in \mathbb{R}$ with $t + \Delta \in [0,1]$,

$$
\begin{aligned}
\sup_{\mathbb{P} \in \mathcal{P}} |h_{\mathbb{P}}(t) - h_{\mathbb{P}}(t + \Delta)| &= \sup_{\mathbb{P} \in \mathcal{P}} \left| \int_{t+(\Delta \wedge 0)}^{t+(\Delta \vee 0)} \phi(F_{\mathbb{P}}^{\leftarrow}(s)) \mathbb{1}_{[0, F_{\mathbb{P}}(0)]}(s)\, ds \right| \\
&\leq \sup_{\mathbb{P} \in \mathcal{P}} \int_0^{|\Delta|} \phi(F_{\mathbb{P}}^{\leftarrow}(s)) \mathbb{1}_{[0, F_{\mathbb{P}}(0)]}(s)\, ds \\
&= \sup_{\mathbb{P} \in \mathcal{P}} h_{\mathbb{P}}(|\Delta|),
\end{aligned}
$$

where we used the fact that $\phi(F_{\mathbb{P}}^{\leftarrow}(\cdot)) \mathbb{1}_{[0, F_{\mathbb{P}}(0)]}(\cdot)$ is nonincreasing on $[0,1]$.

Step 2. Next, we prepare for Step 3. On the one hand, by the uniform equicontinuity of the family $\{h_{\mathbb{P}}\}_{\mathbb{P} \in \mathcal{P}}$ on the compact interval $[0,1]$ (cf. Step 1), we can find for every $\varepsilon > 0$ a finite partition $0 = s_0^{\varepsilon} < s_1^{\varepsilon} < \cdots < s_{k_\varepsilon}^{\varepsilon} = 1$ (being independent of $\mathbb{P}$) such that

$$
\sup_{\mathbb{P} \in \mathcal{P}} \sup_{i=1,\ldots,k_\varepsilon} \int_{s_{i-1}^{\varepsilon}}^{s_i^{\varepsilon}} w_{\mathbb{P}}(s)\, ds \left(= \sup_{\mathbb{P} \in \mathcal{P}} \sup_{i=1,\ldots,k_\varepsilon} (h_{\mathbb{P}}(s_i^{\varepsilon}) - h_{\mathbb{P}}(s_{i-1}^{\varepsilon})) \right) \leq \varepsilon/2.
$$

On the other hand, by assumption (b) we may choose a constant $K_\varepsilon > 0$ such that $\sup_{\mathbb{P} \in \mathcal{P}} \int \phi(z) \mathbb{1}_{\phi(z) \geq K_\varepsilon} \mathbb{P}_1(dz) \leq \varepsilon/2$. Thus, noting

$$
\begin{aligned}
w_{\mathbb{P}}^{\rightarrow}(y) &= \sup\{s \in [0,1] : \phi(F_{\mathbb{P}}^{\leftarrow}(s)) \mathbb{1}_{[0, F_{\mathbb{P}}(0)]}(s) > y\} \\
&\leq \sup\{s \in [0,1] : s \leq F_{\mathbb{P}}(\phi^{\rightarrow}(y)) \wedge F_{\mathbb{P}}(0)\} \\
&\leq F_{\mathbb{P}}(\phi^{\rightarrow}(y))
\end{aligned}
$$

for $y \in (0, \infty)$, and using integration-by-parts (more precisely Theorem 1.15 in [23]), we obtain

$$
\begin{aligned}
\sup_{\mathbb{P} \in \mathcal{P}} \int_{K_\varepsilon}^{\infty} w_{\mathbb{P}}^{\rightarrow}(y)\, dy &\leq \sup_{\mathbb{P} \in \mathcal{P}} \int_{K_\varepsilon}^{\infty} F_{\mathbb{P}}(\phi^{\rightarrow}(y))\, dy \\
&= \sup_{\mathbb{P} \in \mathcal{P}} \left(\int_{(-\infty, \phi^{\rightarrow}(K_\varepsilon)]} \phi(z)\, \mathbb{P}_1(dz) - K_\varepsilon\, F_{\mathbb{P}}(\phi^{\rightarrow}(K_\varepsilon)) \right) \\
&\leq \sup_{\mathbb{P} \in \mathcal{P}} \int_{(-\infty, \phi^{\rightarrow}(K_\varepsilon)]} \phi(z)\, \mathbb{P}_1(dz) \\
&\leq \sup_{\mathbb{P} \in \mathcal{P}} \int_{(-\infty, 0]} \phi(z)\, \mathbb{1}_{\phi(z) \geq K_\varepsilon}\, \mathbb{P}_1(dz) \\
&\leq \varepsilon/2,
\end{aligned}
$$

where $\phi^{\rightarrow}(z) := \sup\{y \in (-\infty, 0] : \phi(y) > z\}$, $z \in [0, \infty)$, denotes the rightcontinuous inverse of the nonincreasing function $\phi : (-\infty, 0] \to [1, \infty)$. Taking into account that the functions $w_{\mathbb{P}}^{\rightarrow}$ take values only in $[0,1]$, we can find a finite partition $0 = y_0^{\varepsilon} < y_1^{\varepsilon} < \cdots < y_{l_\varepsilon - 1}^{\varepsilon} < y_{l_\varepsilon}^{\varepsilon} = \infty$ (being independent of $\mathbb{P}$), with $y_{l_\varepsilon - 1}^{\varepsilon} = K_\varepsilon$, such that

$$
\sup_{\mathbb{P} \in \mathcal{P}} \sup_{i=1,\ldots,l_\varepsilon} \int_{y_{i-1}^{\varepsilon}}^{y_i^{\varepsilon}} w_{\mathbb{P}}^{\rightarrow}(y)\, dy \leq \varepsilon/2,
$$

i.e., in other words,

$$\sup_{\mathbb{P} \in \mathcal{P}} \sup_{i=1,\dots,l_\varepsilon} \int_0^1 \Big((y_i^\varepsilon \wedge w_{\mathbb{P}}(s)) - (y_{i-1}^\varepsilon \wedge w_{\mathbb{P}}(s)) \Big) ds \; \le \; \varepsilon/2.$$

Finally, for every $\mathbb{P} \in \mathcal{P}$, we let $0 = t_{0,\mathbb{P}}^\varepsilon < t_{1,\mathbb{P}}^\varepsilon < \cdots < t_{m_{\varepsilon,\mathbb{P}},\mathbb{P}}^\varepsilon = 1$ be the partition consisting of all points s_i^ε and $w_{\mathbb{P}}^{\rightarrow}(y_i^\varepsilon)$, where $m_{\varepsilon,\mathbb{P}} \le k_\varepsilon + l_\varepsilon$, and $k_\varepsilon + l_\varepsilon =: m_\varepsilon$ is independent of $\mathbb{P}$. For notational simplicity we assume without loss of generality $m_{\varepsilon,\mathbb{P}} = m_\varepsilon$ for all $\mathbb{P} \in \mathcal{P}$.

Step 3. Let $L^1(d\mathbb{I})$ be the space of all Lebesgue integrable functions on $[0,1]$, and $[l,u] := \{f \in L^1(d\mathbb{I}) : l \le f \le u\}$ be the bracket of two functions $l, u \in L^1(d\mathbb{I})$ with $l \le u$ pointwise. For any $\varepsilon > 0$, a bracket $[l,u]$ is called ε-bracket if $\int_0^1 (u-l)\, d\mathbb{I} < \varepsilon$; cf. [34, p. 83]. Using the notation introduced in Step 2, we set for every $\mathbb{P} \in \mathcal{P}$ and $i = 1, \dots, m_\varepsilon$

$$\begin{aligned}
l_{i,\mathbb{P}}^\varepsilon(\cdot) &:= w_{\mathbb{P}}(t_{i,\mathbb{P}}^\varepsilon)\mathbb{1}_{[0,t_{i-1,\mathbb{P}}^\varepsilon]}(\cdot), \\
u_{i,\mathbb{P}}^\varepsilon(\cdot) &:= w_{\mathbb{P}}(t_{i-1,\mathbb{P}}^\varepsilon)\mathbb{1}_{[0,t_{i-1,\mathbb{P}}^\varepsilon]}(\cdot) + w_{\mathbb{P}}(\cdot)\,\mathbb{1}_{(t_{i-1,\mathbb{P}}^\varepsilon, t_{i,\mathbb{P}}^\varepsilon]}(\cdot).
\end{aligned}$$

It follows from the choice of the $t_{i,\mathbb{P}}^\varepsilon$ that, for every $\mathbb{P} \in \mathcal{P}$, $[l_{1,\mathbb{P}}^\varepsilon, u_{1,\mathbb{P}}^\varepsilon], \dots, [l_{m_\varepsilon,\mathbb{P}}^\varepsilon, u_{m_\varepsilon,\mathbb{P}}^\varepsilon]$ provide ε-brackets in $L^1(d\mathbb{I})$ covering the class $\mathcal{E}_{\mathbb{P}} := \{w_{s,\mathbb{P}} : s \in [0,1]\}$ of functions

$$w_{s,\mathbb{P}}(\cdot) := w_{\mathbb{P}}(s)\mathbb{1}_{[0,s]}(\cdot).$$

Step 4. By the usual quantile transformation [31, p. 103], we can find for every $\mathbb{P} \in \mathcal{P}$ a sequence of $U[0,1]$-random variables $U_{1,\mathbb{P}}, U_{2,\mathbb{P}}, \dots$ (possibly on an extension $(\Omega_{\mathbb{P}}, \mathcal{F}_{\mathbb{P}}, \overline{\mathbb{P}})$ of the original probability space $(\Omega, \mathcal{F}, \mathbb{P})$) such that the sequence $(U_{i,\mathbb{P}})$ has the same mixing coefficients (under $\overline{\mathbb{P}}$) as the sequence (X_i) under $\mathbb{P}$ and such that the corresponding empirical distribution function $\widehat{G}_{n,\mathbb{P}}$ satisfies $\widehat{F}_n = \widehat{G}_{n,\mathbb{P}} \circ F_{\mathbb{P}}$ $\overline{\mathbb{P}}$-almost surely. Here we will show as in the proof of Theorem 2.4.1 in [34] that $\overline{\mathbb{P}}$-almost surely

$$\begin{aligned}
&\sup_{y \le 0} |\widehat{F}_n(y) - F_{\mathbb{P}}(y)|\,\phi(y) \\
&\le \max_{i=1,\dots,m_\varepsilon} \max\left\{ \int_0^1 u_{i,\mathbb{P}}^\varepsilon\, d(\widehat{G}_{n,\mathbb{P}} - \mathbb{I})\,;\, \int_0^1 l_{i,\mathbb{P}}^\varepsilon\, d(\mathbb{I} - \widehat{G}_{n,\mathbb{P}}) \right\} + \varepsilon \qquad (17)
\end{aligned}$$

for every $\varepsilon > 0$. Since $F_{\mathbb{P}}^{\leftarrow}(F_{\mathbb{P}}(y)) \le y$ for all $y \in \mathbb{R}$ (cf. [30, p. 113]) and ϕ is nonincreasing on $(-\infty, 0]$, we have $\overline{\mathbb{P}}$-almost surely

$$\begin{aligned}
\sup_{x \le 0} |\widehat{F}_n(y) - F_{\mathbb{P}}(y)|\,\phi(y) &= \sup_{y \le 0} |\widehat{G}_{n,\mathbb{P}}(F_{\mathbb{P}}(y)) - F_{\mathbb{P}}(y)|\,\phi(y) \\
&\le \sup_{y \le 0} |\widehat{G}_{n,\mathbb{P}}(F_{\mathbb{P}}(y)) - F_{\mathbb{P}}(y)|\,\phi(F_{\mathbb{P}}^{\leftarrow}(F_{\mathbb{P}}(y))) \\
&\le \sup_{s \in (0,1)} |\widehat{G}_{n,\mathbb{P}}(s) - s|\,w_{\mathbb{P}}(s) \\
&= \sup_{s \in (0,1)} \left| \int_0^1 w_{s,\mathbb{P}}\, d\widehat{G}_{n,\mathbb{P}} - \int_0^1 w_{s,\mathbb{P}}\, d\mathbb{I} \right|.
\end{aligned}$$

16

So for (17) it suffices to show that

$$\sup_{s \in (0,1)} \left| \int_0^1 w_{s,\mathbb{P}}\, d\widehat{G}_{n,\mathbb{P}} - \int_0^1 w_{s,\mathbb{P}}\, d\mathbb{I} \right|$$

$$\leq \max_{i=1,\ldots,m_\varepsilon} \max \left\{ \int_0^1 u_{i,\mathbb{P}}^\varepsilon\, d(\widehat{G}_{n,\mathbb{P}} - \mathbb{I})\, ; \int_0^1 l_{i,\mathbb{P}}^\varepsilon\, d(\mathbb{I} - \widehat{G}_{n,\mathbb{P}}) \right\} + \varepsilon. \qquad (18)$$

To prove (18), we note that for every $s \in [0,1]$ there is some $i_s = i_s(\mathbb{P}) \in \{1,\ldots,m_\varepsilon\}$ such that $w_{s,\mathbb{P}} \in [l_{i_s,\mathbb{P}}^\varepsilon, u_{i_s,\mathbb{P}}^\varepsilon]$; cf. Step 3. Therefore, since $[l_{i_s,\mathbb{P}}^\varepsilon, u_{i_s,\mathbb{P}}^\varepsilon]$ is an ε-bracket,

$$
\begin{aligned}
\int_0^1 w_{s,\mathbb{P}}\, d\widehat{G}_{n,\mathbb{P}} - \int_0^1 w_{s,\mathbb{P}}\, d\mathbb{I}
&\leq \int_0^1 u_{i_s,\mathbb{P}}^\varepsilon\, d\widehat{G}_{n,\mathbb{P}} - \int_0^1 w_{s,\mathbb{P}}\, d\mathbb{I} \\
&= \int_0^1 u_{i_s,\mathbb{P}}^\varepsilon\, d(\widehat{G}_{n,\mathbb{P}} - \mathbb{I}) + \int_0^1 (u_{i_s,\mathbb{P}}^\varepsilon - w_{s,\mathbb{P}})\, d\mathbb{I} \\
&\leq \int_0^1 u_{i_s,\mathbb{P}}^\varepsilon\, d(\widehat{G}_{n,\mathbb{P}} - \mathbb{I}) + \int_0^1 (u_{i_s,\mathbb{P}}^\varepsilon - l_{i_s,\mathbb{P}}^\varepsilon)\, d\mathbb{I} \\
&\leq \max_{i=1,\ldots,m_\varepsilon} \int_0^1 u_{i,\mathbb{P}}^\varepsilon\, d(\widehat{G}_{n,\mathbb{P}} - \mathbb{I}) + \varepsilon.
\end{aligned}
$$

Analogously we obtain

$$\int_0^1 w_{s,\mathbb{P}}\, d\widehat{G}_{n,\mathbb{P}} - \int_0^1 w_{s,\mathbb{P}}\, d\mathbb{I} \geq -\left(\max_{i=1,\ldots,m_\varepsilon} \int_0^1 l_{i,\mathbb{P}}^\varepsilon\, d(\mathbb{I} - \widehat{G}_{n,\mathbb{P}}) + \varepsilon \right).$$

That is, (18) and therefore (17) hold true.

Step 5. Because of (17), for (16) to be true it suffices to show that for every $\delta > 0$

$$\lim_{n\to\infty} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\Big[\max_{i=1,\ldots,m_\varepsilon} \max \left\{ \int_0^1 u_{i,\mathbb{P}}^\varepsilon\, d(\widehat{G}_{n,\mathbb{P}} - \mathbb{I})\, ; \int_0^1 l_{i,\mathbb{P}}^\varepsilon\, d(\mathbb{I} - \widehat{G}_{n,\mathbb{P}}) \right\} \geq \delta/2 \Big] = 0 \quad (19)$$

with $\varepsilon = \varepsilon(\delta) := \delta/2$. For (19) to be true it suffices to show that

$$\lim_{n\to\infty} \sup_{\mathbb{P}\in\mathcal{P}} \overline{\mathbb{P}}\Big[\left| \int_0^1 l_{i,\mathbb{P}}^\varepsilon\, d(\mathbb{I} - \widehat{G}_{n,\mathbb{P}}) \right| \geq \delta/2 \Big] = 0, \qquad (20)$$

$$\lim_{n\to\infty} \sup_{\mathbb{P}\in\mathcal{P}} \overline{\mathbb{P}}\Big[\left| \int_0^1 u_{i,\mathbb{P}}^\varepsilon\, d(\widehat{G}_{n,\mathbb{P}} - \mathbb{I}) \right| \geq \delta/2 \Big] = 0, \qquad (21)$$

for every $i = 1,\ldots,m_\varepsilon$, because $\overline{\mathbb{P}}$ is subadditive. We will show only (21). Assertion (20) can be shown analogously. Since

$$
\begin{aligned}
&\int_0^1 u_{i,\mathbb{P}}^\varepsilon\, d(\widehat{G}_{n,\mathbb{P}} - \mathbb{I}) \\
&= \frac{1}{n} \sum_{j=1}^n \Big(w_{\mathbb{P}}(t_{i-1,\mathbb{P}}^\varepsilon)\, \mathbb{1}_{[0,t_{i-1,\mathbb{P}}^\varepsilon]}(U_{j,\mathbb{P}}) - \mathbb{E}_{\overline{\mathbb{P}}}\Big[w_{\mathbb{P}}(t_{i-1,\mathbb{P}}^\varepsilon)\, \mathbb{1}_{[0,t_{i-1,\mathbb{P}}^\varepsilon]}(U_{1,\mathbb{P}}) \Big] \Big) \\
&\quad + \frac{1}{n} \sum_{j=1}^n \Big(w_{\mathbb{P}}(U_{j,\mathbb{P}})\, \mathbb{1}_{(t_{i-1,\mathbb{P}}^\varepsilon, t_{i,\mathbb{P}}^\varepsilon]}(U_{j,\mathbb{P}}) - \mathbb{E}_{\overline{\mathbb{P}}}\Big[w_{\mathbb{P}}(U_{1,\mathbb{P}})\, \mathbb{1}_{(t_{i-1,\mathbb{P}}^\varepsilon, t_{i,\mathbb{P}}^\varepsilon]}(U_{1,\mathbb{P}}) \Big] \Big),
\end{aligned}
$$

17

for (21) to be true it suffices to show that for every fixed $i \in \{1, \ldots, m_\varepsilon\}$

$$\lim_{n \to \infty} \sup_{\mathbb{P} \in \mathcal{P}} \overline{\mathbb{P}}\Big[\Big| \frac{1}{n} \sum_{j=1}^{n} \Big(w_{\mathbb{P}}(t_{i-1,\mathbb{P}}^\varepsilon) \mathbb{1}_{[0, t_{i-1,\mathbb{P}}^\varepsilon]}(U_{j,\mathbb{P}})$$
$$- \mathbb{E}_{\mathbb{P}}\Big[w_{\mathbb{P}}(t_{i-1,\mathbb{P}}^\varepsilon) \mathbb{1}_{[0, t_{i-1,\mathbb{P}}^\varepsilon]}(U_{1,\mathbb{P}})\Big]\Big) \Big| \geq \delta/4 \Big] = 0, \tag{22}$$

$$\lim_{n \to \infty} \sup_{\mathbb{P} \in \mathcal{P}} \overline{\mathbb{P}}\Big[\Big| \frac{1}{n} \sum_{j=1}^{n} \Big(w_{\mathbb{P}}(U_{j,\mathbb{P}}) \mathbb{1}_{(t_{i-1,\mathbb{P}}^\varepsilon, t_{i,\mathbb{P}}^\varepsilon]}(U_{j,\mathbb{P}})$$
$$- \mathbb{E}_{\mathbb{P}}\Big[w_{\mathbb{P}}(U_{1,\mathbb{P}}) \mathbb{1}_{(t_{i-1,\mathbb{P}}^\varepsilon, t_{i,\mathbb{P}}^\varepsilon]}(U_{1,\mathbb{P}})\Big]\Big) \Big| \geq \delta/4 \Big] = 0. \tag{23}$$

We will show only (23). Assertion (22) can be shown similarly, noting that the inequality $w_{\mathbb{P}}(t_{i-1,\mathbb{P}}^\varepsilon) \mathbb{1}_{[0, t_{i-1,\mathbb{P}}^\varepsilon]}(U_{1,\mathbb{P}}) \leq w_{\mathbb{P}}(U_{1,\mathbb{P}}) \mathbb{1}_{[0, t_{i-1,\mathbb{P}}^\varepsilon]}(U_{1,\mathbb{P}})$ holds since $w_{\mathbb{P}}$ is nonincreasing.

Corollary A.2 ensures (23) if we can show that the conditions (26) and (27) in the corollary hold for $\Pi := \mathcal{P}$, $(\Omega_\pi, \mathcal{F}_\pi, \mathbb{P}_\pi) := (\Omega_{\mathbb{P}}, \mathcal{F}_{\mathbb{P}}, \overline{\mathbb{P}})$ and $\xi_{j,\pi}^{(i)} := w_{\mathbb{P}}(U_{j,\mathbb{P}}) \mathbb{1}_{(t_{i-1,\mathbb{P}}^\varepsilon, t_{i,\mathbb{P}}^\varepsilon]}(U_{j,\mathbb{P}})$, $j \in \mathbb{N}$, for every fixed $i \in \{1, \ldots, m_\varepsilon\}$. Condition (26) follows from

$$\limsup_{K \to \infty} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\overline{\mathbb{P}}}\Big[|\xi_{1,\mathbb{P}}^{(i)}| \, \mathbb{1}_{|\xi_{1,\mathbb{P}}^{(i)}| \geq K}\Big] \leq \limsup_{K \to \infty} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\overline{\mathbb{P}}}\Big[|w_{\mathbb{P}}(U_{1,\mathbb{P}})| \, \mathbb{1}_{|w_{\mathbb{P}}(U_{1,\mathbb{P}})| \geq K}\Big]$$

$$\leq \limsup_{K \to \infty} \sup_{\mathbb{P} \in \mathcal{P}} \int_0^1 |\phi(F_{\mathbb{P}}^\leftarrow(t))| \, \mathbb{1}_{|\phi(F_{\mathbb{P}}^\leftarrow(t))| \geq K} \, dt$$

$$= \limsup_{K \to \infty} \sup_{\mathbb{P} \in \mathcal{P}} \int \phi(y) \mathbb{1}_{\phi(y) \geq K} \, \mathbb{P}_1(dy)$$

and assumption (b). Condition (27) is an immediate consequence of assumption (c), Remark A.3 and the fact that the sequence $(U_{j,\mathbb{P}})$ has the same mixing coefficients under $\overline{\mathbb{P}}$ as the sequence (X_i) under $\mathbb{P}$. This completes the proof of Theorem 3.1.

6. Proof of Theorem 3.9

Let $\delta > 0$ be arbitrary but fixed. Choose $k_\delta \in \mathbb{N}$ such that $\sum_{k=k_\delta+1}^{\infty} 2^{-k} < \delta/3$, and notice that (9) holds if we can show that the following two conditions hold

$$\lim_{n \to \infty} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\Big[\Big| \int f_k \, d\widehat{m}_n - \int f_k \, d\mathbb{P}_1 \Big| \geq \frac{\delta}{3k_\delta} \Big] = 0, \qquad k = 1, \ldots, k_\delta, \tag{24}$$

$$\lim_{n \to \infty} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\Big[\Big| \int \psi \, d\widehat{m}_n - \int \psi \, d\mathbb{P}_1 \Big| \geq \frac{\delta}{3} \Big] = 0. \tag{25}$$

To prove (25), we note that the left-hand side in (25) can be written as

$$\lim_{n \to \infty} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\Big[\Big| \frac{1}{n} \sum_{j=i}^{n} \psi(X_i) - \mathbb{E}_{\mathbb{P}}[\psi(X_1)] \Big| \geq \frac{\delta}{3} \Big].$$

The latter is zero by Corollary A.2 with $\Pi := \mathcal{P}$, $(\Omega_\pi, \mathcal{F}_\pi, \mathbb{P}_\pi) := (\Omega, \mathcal{F}, \mathbb{P})$ and the random variables $\xi_{i,\pi} := \psi(X_i)$, $i \in \mathbb{N}$. Indeed: Assumption (26) of the corollary holds by our assumption (b) and the identity $\mathbb{E}_\pi[|\xi_{1,\pi}|\mathbb{1}_{|\xi_{1,\pi}| \geq K}] = \int \psi \mathbb{1}_{\psi \geq K} \, d\mathbb{P}_1$. Assumption (27) of the corollary holds by the fact that, under every $\mathbb{P} \in \mathcal{P}$, the sequence $(\psi(X_i))$ has the same mixing coefficients as the sequence (X_i) and by our assumption (c) along with Remark A.3.

Assertion (24) can be proven in the same line, noting that $\mathcal{P}$ also satisfies condition (b) with ψ replaced by f_k for any $k = 1, \ldots, k_\delta$; recall that each f_k has compact support. This completes the proof of Theorem 3.9.

A. Uniform weak LLN for strongly mixing random variables

Let $(\xi_i) = (\xi_i)_{i \in \mathbb{N}}$ be a sequence of random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. According to Rosenblatt [29], the sequence (ξ_i) is said to be *strongly mixing* (or α-*mixing*) if the nth mixing coefficient $\alpha_\mathbb{P}(n) := \sup_{k \geq 1} \sup_{A,B} |\mathbb{P}[A \cap B] - \mathbb{P}[A]\mathbb{P}[B]|$ converges to zero as $n \to \infty$, where the second supremum ranges over all $A \in \sigma(\xi_1, \ldots, \xi_k)$ and $B \in \sigma(\xi_{k+n}, \xi_{k+n+1}, \ldots)$. Recall from [4, Inequality (1.9)] that $\alpha_\mathbb{P}(n) \leq 1/4$ for all $n \in \mathbb{N}$, and use the convention $\alpha_\mathbb{P}(0) := 1/4$. For an overview on mixing conditions see [4, 13]. The following result is a consequence of Theorem 4 and Lemma 1 in [28]; cf. (5.1) on page 936 in [28].

Theorem A.1 (Rio) *Let $\xi_1, \xi_2, \ldots$ be identically distributed with $\mathbb{E}[\xi_1^2] < \infty$, and $\alpha_\mathbb{P}(n)$ be defined as above. Let G be the distribution function of $|\xi_1|$, and set $\overline{G} := 1 - G$. Then, for every $x > 0$ and $n \in \mathbb{N}$,*

$$\mathbb{P}\left[\sup_{k=1,\ldots,n} \left| \sum_{i=1}^{k} (\xi_i - \mathbb{E}[\xi_1]) \right| \geq 2x \right] \leq \frac{16}{x^2} n \sum_{j=0}^{n-1} \int_0^{2\alpha_\mathbb{P}(j)} \overline{G}^{\rightarrow}(t)^2 \, dt.$$

From Theorem A.1 we can even derive the following uniform weak LLN for strongly mixing random variables. In the special case of independent random variables the corollary is already known from [9]. In fact, [9] provides even a uniform strong LLN.

Corollary A.2 *Let $\Pi \neq \emptyset$ be an arbitrary index set. Further, for every $\pi \in \Pi$, let $(\Omega_\pi, \mathcal{F}_\pi, \mathbb{P}_\pi)$ be a probability space and $\xi_{1,\pi}, \xi_{2,\pi}, \ldots$ be a sequence of random variables on $(\Omega_\pi, \mathcal{F}_\pi, \mathbb{P}_\pi)$ being identically distributed and strongly mixing with mixing coefficients $(\alpha_\pi(n)) := (\alpha_{\mathbb{P}_\pi}(n))$. Further suppose that the following two conditions hold*

$$\lim_{K \to \infty} \sup_{\pi \in \Pi} \mathbb{E}_\pi[|\xi_{1,\pi}|\mathbb{1}_{|\xi_{1,\pi}| \geq K}] = 0, \tag{26}$$

$$\lim_{n \to \infty} \sup_{\pi \in \Pi} \frac{1}{n} \sum_{j=0}^{n-1} \alpha_\pi(j) = 0. \tag{27}$$

Then, for every $\delta > 0$,

$$\lim_{n \to \infty} \sup_{\pi \in \Pi} \mathbb{P}_\pi \Big[\Big| \frac{1}{n} \sum_{i=1}^n \xi_{i,\pi} - \mathbb{E}_\pi[\xi_{1,\pi}] \Big| \geq \delta \Big] = 0. \tag{28}$$

Remark A.3 By the Toeplitz lemma, (27) holds if $\lim_{n \to \infty} \sup_{\pi \in \Pi} \alpha_\pi(n) = 0$. $\diamond$

Proof (of Corollary A.2) We will use a truncation argument. Let $K > 0$ be a constant (to be specified later on) and $\xi_{i,\pi}^K := \xi_{i,\pi} \mathbb{1}_{|\xi_{i,\pi}| \leq K}$ be the K-truncation of $\xi_{i,\pi}$. Using the decomposition $\xi_{i,\pi} = \xi_{i,\pi}^K + \xi_{i,\pi} \mathbb{1}_{|\xi_{i,\pi}| > K}$ and the triangular inequality, we obtain

$$
\begin{aligned}
\mathbb{P}_\pi \Big[\Big| \frac{1}{n} \sum_{i=1}^n \xi_{i,\pi} - \mathbb{E}_\pi[\xi_{1,\pi}] \Big| \geq \delta \Big] \quad & \leq \quad \mathbb{P}_\pi \Big[\Big| \frac{1}{n} \sum_{i=1}^n \xi_{i,\pi}^K - \mathbb{E}_\pi[\xi_{1,\pi}^K] \Big| \geq \delta/3 \Big] \\
& \quad + \mathbb{P}_\pi \Big[\Big| \frac{1}{n} \sum_{i=1}^n \xi_{i,\pi} \mathbb{1}_{|\xi_{i,\pi}| > K} \Big| \geq \delta/3 \Big] \\
& \quad + \mathbb{P}_\pi \Big[\mathbb{E}_\pi[|\xi_{1,\pi}| \mathbb{1}_{|\xi_{1,\pi}| > K}] \geq \delta/3 \Big] \\
& =: \quad S_1(\delta, n, K, \pi) + S_2(\delta, n, K, \pi) + S_3(\delta, K, \pi).
\end{aligned}
$$

By Markov's inequality, $S_2(\delta, n, K, \pi)$ is bounded above by $3\delta^{-1} \sup_{\pi \in \Pi} \mathbb{E}_\pi[|\xi_{1,\pi}| \mathbb{1}_{|\xi_{1,\pi}| \geq K}]$. So, for given $\varepsilon > 0$, one can choose $K = K_{\varepsilon, \delta}$ in such a way that

$$\sup_{\pi \in \Pi} S_2(\delta, n, K, \pi) \leq \varepsilon/2 \qquad \forall\, n \in \mathbb{N},$$

because we assumed (26). By (26), we may and do also assume that $K = K_{\varepsilon, \delta}$ is chosen such that

$$\sup_{\pi \in \Pi} S_3(\delta, K, \pi) = 0.$$

Choosing $x := n\delta/6$ in Theorem A.1, we further obtain

$$
\begin{aligned}
\sup_{\pi \in \Pi} S_1(\delta, n, K, \pi) \quad & \leq \quad \frac{576}{\delta^2} \sup_{\pi \in \Pi} \frac{1}{n} \sum_{j=0}^{n-1} \int_0^{2\alpha_\pi(j)} \overline{G}_{K,\pi}^{\rightarrow}(t)^2 \, dt \\
& \leq \quad \frac{1152\, K^2}{\delta^2} \sup_{\pi \in \Pi} \frac{1}{n} \sum_{j=0}^{n-1} \alpha_\pi(j),
\end{aligned}
$$

where $G_{K,\pi}$ denotes the distribution function of $|\xi_{1,\pi}^K|$. So, in view of (27), we can find some $n_\varepsilon \in \mathbb{N}$ such that

$$\sup_{\pi \in \Pi} S_1(\delta, n, K, \pi) \leq \varepsilon/2 \qquad \forall\, n \geq n_\varepsilon.$$

Thus, (28) holds. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

B. Strong mixing of linear processes

Let $(Z_s)_{s \in \mathbb{Z}}$ be i.i.d. random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and define the linear process $X_t := \sum_{s=0}^{\infty} a_s Z_{t-s}$, $t \in \mathbb{N}$, for any real sequence $(a_s)_{s \in \mathbb{N}_0}$ for which the latter series is $\mathbb{P}$-almost surely well defined for every $t \in \mathbb{N}$. Moreover, let $\alpha_{\mathbb{P}}(n)$ be the nth strong mixing coefficient of the sequence (X_t) under $\mathbb{P}$ as defined at the beginning of Section A. The following criterion for (X_t) to be strongly mixing is an immediate consequence of Lemmas 2.1 and 2.2 in [26]; notice that we will assume $a_0 = 1$.

Theorem B.1 *Let $a_0 = 1$, and assume that the following assertions hold:*

(a) *The distribution of Z_1 admits a Lebesgue density f for which $\int |f(y+h) - f(y)| \, dy < M|h|$ for all $h \in \mathbb{R}$ and some constant $M > 0$.*

(b) $\mathbb{E}[|Z_1|] < \infty$.

(c) $\sum_{s=0}^{\infty} a_s z^s \neq 0$ for all z with $|z| \leq 1$.

(d) $\sum_{u=1}^{\infty} \sum_{s=u}^{\infty} |a_s| < \infty$.

Then, for every $n \in \mathbb{N}$,

$$\alpha_{\mathbb{P}}(n) \leq \left(2M \, \mathbb{E}[|Z_1|] \sum_{s=0}^{\infty} |b_s| \right) \sum_{u=n}^{\infty} \sum_{s=u}^{\infty} |a_s|, \tag{29}$$

where b_s is the coefficient of z^s in the power series expansion of $z \mapsto 1/\sum_{s=0}^{\infty} a_s z^s$. In particular, (X_t) is strongly mixing.

It can be seen from Lemma 2.1 in [26] that the right-hand side in (29) provides an upper bound even for the β-mixing coefficient, i.e. for the mixing coefficient in the context of absolute regularity. Notice that (d) implies $\sum_{s=0}^{\infty} |a_s| < \infty$ which, together with (c) and Wiener's theorem (cf. [35, p. 91] or [37, p. 301]), ensures that $\sum_{s=0}^{\infty} |b_s| < \infty$. Further conditions for a linear process to be strongly mixing can be found in [15, 26, 36].

Remark B.2 (i) Condition (d) is satisfied if $|a_s| \leq Cs^{-\gamma}$ for all $s \in \mathbb{N}_0$ and some constants $C > 0$, $\gamma > 2$. In this case, $\sum_{u=n}^{\infty} \sum_{s=u}^{\infty} |a_s| \leq C_\gamma n^{2-\gamma}$ holds for all $n \in \mathbb{N}$.

(ii) Condition (d) is also satisfied if $|a_s| \leq Cq^s$ for all $s \in \mathbb{N}_0$ and some constants $C > 0$, $q \in (0, 1)$. In this case, $\sum_{u=n}^{\infty} \sum_{s=u}^{\infty} |a_s| \leq C(1-q)^{-2} q^n$ holds for all $n \in \mathbb{N}$. $\diamond$

Example B.3 If $a_s = aq^s$, $s \geq 1$, with $a \neq 0$ and $|q| < 1$, then we have for $|z| \leq 1$

$$\sum_{s=0}^{\infty} a_s z^s = (a_0 - a) + \frac{a}{1 - qz} = \frac{a_0 - \{(a_0 - a)q\}z}{1 - qz} \, .$$

From here one easily derives that $1/\sum_{s=0}^{\infty} a_s z^s$ admits the representation $\sum_{s=0}^{\infty} b_s z^s$ with $b_0 = 1/a_0$ and $b_s = -a(a_0 - a)^{s-1} a_0^{-(s+1)} q^s$, $s \geq 1$, using the convention $0^0 := 1$.

If specifically $a_0 = 1$, $a = (\phi_1 + \theta_1)/\phi_1$ and $q = \phi_1$ for real numbers ϕ_1 and θ_1 satisfying $0 < |\phi_1| < 1$ and $|\theta_1| < 1$, then $\sum_{s=0}^{\infty} |b_s| = 1 + |\phi_1 + \theta_1|/(1 - |\theta_1|)$. $\diamond$

Example B.4 *(ARMA process)* To illustrate conditions (c)–(d) in Theorem B.1, let us consider an ARMA(p,q) process $X_t = \sum_{s=1}^{p} \phi_s X_{t-s} + \sum_{s=0}^{q} \theta_s Z_{t-s}$ with $\theta_0 = 1$ and square-integrable and centered i.i.d. innovations $(Z_s)_{s \in \mathbb{Z}}$. Define the characteristic polynomials $\phi(z) := 1 - \sum_{s=1}^{p} \phi_s z^s$ and $\theta(z) := \sum_{s=0}^{q} \theta_s z^s$, and assume that $\phi(\cdot)$ and $\theta(\cdot)$ have no common zeros. Further assume that (X_t) is both causal and invertible. By the causality we have that $\phi(z) \neq 0$ for all complex z with $|z| \leq 1$, and that X admits the MA(∞) representation $X_t = \sum_{s=0}^{\infty} a_s Z_{t-s}$ with the coefficients a_s determined by

$$\frac{\theta(z)}{\phi(z)} = a(z) := \sum_{s=0}^{\infty} a_s z^s; \tag{30}$$

see [6, Theorem 3.1.1]. By the invertibility we have in addition that $\theta(z) \neq 0$, and hence $a(z) \neq 0$, for all complex z with $|z| \leq 1$; see [6, Theorem 3.1.2]. That is, under the imposed assumptions the ARMA process can be regarded as a linear process satisfying condition (c) in Theorem B.1.

If specifically $p = q = 1$, $|\phi_1|, |\theta_1| < 1$, and $\phi_1 \neq \theta_1$, then the coefficients a_s determined by (30) read as $a_0 = 1$ and $a_s = (\phi_1 + \theta_1)\phi_1^{s-1}$ $(= ((\phi_1 + \theta_1)/\phi_1)\phi_1^s$ if $\phi_1 \neq 0)$, $s \geq 1$. In this case the ARMA$(1,1)$ process can be seen as a linear process which satisfies not only condition (c) but also condition (d) of Theorem B.1. By part (ii) of Remark B.2 we have in particular $\sum_{u=n}^{\infty} \sum_{s=u}^{\infty} |a_s| = (|\phi_1 + \theta_1|/(1 - |\phi_1|)^2) |\phi_1|^{n-1}$ for all $n \in \mathbb{N}$. Moreover, Example B.3 yields $\sum_{s=0}^{\infty} |b_s| = 1 + |\phi_1 + \theta_1|/(1 - |\theta_1|)$, where b_s is the coefficient of z^s in the power series expansion of $z \mapsto 1/\sum_{s=0}^{\infty} a_s z^s$. $\diamond$

Acknowledgement

References

[1] Bauer, H. (2001). *Measure and integration theory.* de Gruyter, Berlin.

[2] Beutner, E. and Zähle, H. (2013). Continuous mapping approach to the asymptotics of U- and V-statistics. *Bernoulli*, fothcoming.

[3] Boente, G., Fraiman, R. and Yohai, V.J. (1987). Qualitative robustness for stochastic processes. *Annals of Statistics*, 15, 1293–1312.

[4] Bradley, R.C. (2005). Basic properties of strong mixing conditions. A survey and some open problems. *Probability Surveys*, 2, 107–144.

[5] Bradley, R.C. (2007). *Introduction to strong mixing conditions, Volume 1*. Kendrick Press, Heber City, Utah.

[6] Brockwell, P.J. and Davis, R.A. (2006). *Time series: theory and methods*, Springer-Verlag, New York.

[7] Bustos, O.H. (1981). Qualitative robustness for general processes. *Informes de Matematica, Serie B-002/81*, Institute de Matematica Pura e Aplicada, Rio de Janeiro.

[8] Christmann, A., Salibían–Barrera, M., and Van Aelst, S. (2011). On the stability of bootstrap estimators. *Submitted* (arXiv:1111.1876).

[9] Chung, K.L. (1951). The strong law of large numbers. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley and Los Angeles, 341–352.

[10] Cox, D. (1981). Metrics on stochastic processes and qualitative robustness. Technical Report 3, *Department of Statistics, University of Washington*.

[11] Cuevas, A. (1988). Qualitative robustness in abstract inference. *Journal of Statistical Planning and Inference*, 18, 277–289.

[12] Cuevas, A. and Romo, R. (1993). On robust properties of bootstrap approximations. *Journal of Statistical Planning and Inference*, 37, 181–191.

[13] Doukhan, P. (1994). Mixing: properties and examples. *Lecture Notes in Statistics*, Springer-Verlag, New York.

[14] Föllmer, H. and Schied, A. (2004). *Stochastic finance. An introduction in discrete time*. de Gruyter, Berlin.

[15] Gorodetskii, V.V. (1977). On the strong mixing property for linear processes. *Theory of Probability and its Applications*, 22, 411–413.

[16] Hampel F.R. (1968). *Contributions to the theory of robust estimation*. Ph.D. Thesis, University of California, Berkeley.

[17] Hampel, F.R. (1971). A general qualitative definition of robustness. *Annals of Mathematical Statistics*, 42, 1887–1896.

[18] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust statistics – the approach based on influence functions*, Wiley, New York.

[19] Huber, P.J. (1981). *Robust statistics*. Wiley, New York.

[20] Huber, P.J. and Ronchetti E.M. (2009). *Robust statistics*. Wiley, New York.

[21] Krätschmer, V., Schied, A. and Zähle, H. (2012). Qualitative and infinitesimal robustness of tail-dependent statistical functionals. *Journal of Multivariate Analysis*, 103, 35–47.

[22] Krätschmer, V., Schied, A. and Zähle, H. (2012). Comparative and qualitative robustness for law-invariant risk measures. *Submitted* (arXiv:1204.2458).

[23] Mattila, P. (1995). *Geometry of sets and measures in Euclidean spaces*. Cambridge University Press, Cambridge.

[24] Mizera, I. (2010). Qualitative robustness and weak continuity: the extreme unction. *In:* J. Antoch, M. Huškova, P.K. Sen (Eds.), IMS Collections Festschrift in honor of Professor Jana Jurečková: *Nonparametrics and robustness in modern statistical inference and time series analysis*. Institute of Mathematical Statistics, 169–181.

[25] Papantoni-Kazakos, P. and Gray, R.M. (1979). Robustness of estimators on stationary observations. *Annals of Probability*, 7, 989–1002.

[26] Pham, T.D. and Tran, L.T. (1985). Some mixing properties of time series models. *Stochastic Processes and their Applications*, 19, 297–303.

[27] Rieder, H. (1994). *Robust asymptotic statistics*. Springer-Verlag, New York.

[28] Rio, E. (1995). A maximal inequality and dependent Marcinkiewicz-Zygmund strong laws. *Annals of Probability*, 23, 918–937.

[29] Rosenblatt, M. (1956). A central limit theorem and a mixing condition. *Proceedings of the National Academy of Sciences of the USA*, 42, 412–413.

[30] Shorack, G.R (2000). *Probability for statisticians*. Springer, New York.

[31] Shorack, G.R. and Wellner, J.A. (1986). *Empirical processes with applications to statistics*. Wiley, New York.

[32] Strassen, V. (1965). The existence of probability measures with given marignals. *Annals of Mathematical Statistics*, 36, 423–439.

[33] van der Vaart, A.W. (1998). *Asymptotic statistics*. Cambridge University Press, Cambridge.

[34] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes.* Springer, New York.

[35] Wiener, N. (1933). *The Fourier integral and certain of its applications.* Cambridge University Press, Cambridge.

[36] Withers, C.S. (1981). Conditions for linear processes to be strong-mixing. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57, 477–480.

[37] Yosida, K. (1995). *Functional analysis.* Springer-Verlag, Berlin.

[38] Zähle, H. (2012). Marcinkiewicz-Zygmund strong laws for statistical functionals and empirical distribution functions. *Statistics*, forthcoming (arXiv:1301.0726).